数理解析研究所講究録
741巻 1991年 24-30

Stochastic games with constraints

新潟大・理学部　田中謙輔　(Kensuke Tanaka)
ハルビン師範大　劉　兆華　(Zhaohua Liu)

制約条件をもつ２人零和確率ゲームを次の様に１０個の要素で与える事にする：

$$(X,\ Y,\ A,\ B,\ q_1,\ q_2,\ r,\ g^1,\ g^2,\ \beta),$$

ただし，X,Y はそれぞれ player 1, player 2 の状態空間，直積 XY はゲームの状態空間，A,B はそれぞれ player 1, player 2 の行動空間，$q_1(C \mid x,a)$ は $(x,a) \in XA$ が与えられたときの状態空間 X 上の Borel set C の確率，$q_2(D \mid y,b)$ は $(y,b) \in YB$ が与えられたときの状態空間 Y 上の Borel set D の確率，$p(CD \mid (x,a),(y,b))=q_1(C \mid x,a)q_2(D \mid y,b)$ は $((x,a),(y,b)) \in XAYB$ が与えられたときの XY 上の Borel set CD へのゲームの推移確率，$r((x,a),(y,b))$ は実関数：$XAYB \to R$ で player 1 の 1-段階の損失関数で $-r$ は player 2 の 1-段階の損失関数，$g^1=(g^1_1(x,a), g^1_2(x,a),\cdots,g^1_n(x,a))$ はベクトル値関数：$XA \to R^n$ で player 1 の 1-段階の cost 関数，$g^2=(g^2_1(y,b),g^2_2(y,b),\cdots,g^2_m(y,b))$

はベクトル値関数：$YB \to R^m$ で player 2 の 1-段階の cost 関数，β は割引因子で $0 < \beta < 1$ とする．

ここで，player 1 の strategy $\pi=(\pi_1,\pi_2,\cdots,\pi_t,\cdots)$ の各要素 π_t は $H_t=H_{t-1}(AX)$ ($H_1=X$), $t \geqq 2$, が与えられたときのA上の条件付き確率である．特に，π の全ての要素 π_t が時刻 t に関係しないとき，この strategy は定常と呼ばれる．player 2 の strategy $\sigma=(\sigma_1,\sigma_2,\cdots,\sigma_t,\cdots)$ の各要素 σ_t は $H_t=H_{t-1}(BY)$ ($H_1=Y$), $t \geqq 2$, が与えられたときの B 上の条件付き確率である． stategy σ に対しても同様に常定性が与えられる．

このとき，player 1 の總期待割引損失は次の様に与えられる：

$$I(\pi,\sigma) = \sum_{t=1}^{\infty} \beta^{t-1} E_{\pi,\sigma}[r((x_t,a_t),(y_t,b_t))].$$

更に，player 1 の總期待割引cost は次の様に与えられる：

$$G^1(\pi)= \sum_{t=1}^{\infty} \beta^{t-1} E_{\pi}[g^1(x_t,a_t)] \in R^n.$$

亦，player 2 の總期待割引cost は次の様に与えられる：

$$G^2(\sigma)= \sum_{t=1}^{\infty} \beta^{t-1} E_{\sigma}[g^2(y_t,b_t)] \in R^m.$$

次に，$(u,v) \in R^n_+ \times R^m_+$ を用いて，$I(\pi,\sigma)$ を (u,v) の線形関数 $<u,G^1(\pi)>+<v,G^2(\sigma)>$ で perturb したゲームを考察するために次の様な２個の記号を導入することにする：

$$V^+(u,v) = \inf_{\pi} \sup_{\sigma} [I(\pi,\sigma) + < u, G^1(\pi) > - < v, G^2(\sigma) >],$$

と

$$V^-(u,v) = \sup_{\sigma} \inf_{\pi} [I(\pi,\sigma) + < u, G^1(\pi) > - < v, G^2(\sigma) >].$$

定理１　σ_0 が $(\bar{u},\bar{v}) \in R^n_+ \times R^m_+$ の線形関数で perturb した確率ゲームにおける player 2 の max-inf strategy とするとき，次の事が成立する：

$$-G^2(\sigma_0) \in \partial V^+(\bar{u},\cdot)(\bar{v})$$

で

$$\inf_{v \in R^m_+} [V^+(\bar{u},v) + < v, G^2(\sigma_0) >] = \inf_{\pi} [I(\pi,\sigma_0) + < \bar{u}, G^1(\pi) >].$$

ただし，$\partial V^+(\bar{u},\cdot)(\bar{v})$ は v での $V^+(\bar{u},\cdot)$ の subgradient の集合を表す記号で，全ての $z \in \partial V^+(\bar{u},\ )(\bar{v})$ に対して 不等式

$$V^+(\bar{u},v) - V^+(\bar{u},\bar{v}) \geqq <v - \bar{v}, z>$$

が成立している．

定理２　π_0 が $(\bar{u},\bar{v}) \in R^n_+ \times R^m_+$ の線形関数で perturb した確率ゲームにおける player 2 の mini-sup strategy とする

とき，次の事が成立する：

$$G^1(\pi_0) \in \partial V^-(\cdot,\overline{v})(\overline{u})$$

で

$$\sup_{u \in R^n_+} [V^-(u,\overline{v}) - \langle u, G^1(\pi_0) \rangle] = \sup_{\sigma} [I(\pi_0,\sigma) - \langle \overline{v}, G^2(\sigma) \rangle]$$

ただし，ここでは全ての $y \in \partial V^-(\cdot,\overline{v})(\overline{u})$ に対して不等式

$$V^-(u,\overline{v}) - V^-(\overline{u},\overline{v}) \leqq \langle u - \overline{u}, y \rangle$$

が成立している．

定理３　X,Y はそれぞれ Polish space の Borel set, A,B はそれぞれ Polish space の compmct set，推移確率 $p=q_1q_2$ を構成する q_1，q_2 は $(x,y) \in XY$ が与えられたとき，それぞれ A，B 上で連続，r は XAYB 上で連続で有界，g^1 は XA 上で連続で $\|g^1\|$ は有界，g^2 は YB 上で連続で $\|g^2\|$ は有界と仮定すると，$(\overline{u},\overline{v})$ の線形関数で perturb した確率ゲームに対して player 1 と player 2 の定常な strategy の鞍部点(π_0,σ_0)とゲームの値：

$$V^+(\overline{u},\overline{v}) = V^-(\overline{u},\overline{v}) = V(\overline{u},\overline{v})$$

が存在する．ただし，

$$V(\bar{u},\bar{v})=I(\pi_0,\sigma_0)+<\bar{u},G^1(\pi_0)>-<\bar{v},G^2(\sigma_0)>.$$

定理４　定理３の条件のもとで，次の事が成立する．

$$V(\bar{u},\bar{v})+<\bar{v},G^2(\sigma_0)> \leqq \sup_{u\in R^n_+}\inf_{\pi}[I(\pi,\sigma_0)+<u,G^1(\pi)>]$$

$$=\inf_{G^1(\pi)\leqq\theta} I(\pi,\sigma_0)$$

で

$$V(\bar{u},\bar{v})-<\bar{u},G^1(\pi_0)> \geqq \inf_{v\in R^m_+}\sup_{\sigma}[I(\pi_0,\sigma)-<v,G^2(\sigma)>]$$

$$=\sup_{G^2(\sigma)\leqq\theta} I(\pi_0,\sigma).$$

定理５　(π_0,σ_0) が制約条件 $G_1(\pi)\leqq\theta$, $G^2(\sigma)\leqq\theta$ をもつ primal game の鞍部点，即ち

$$\inf_{G^1(\pi)\leqq\theta} I(\pi,\sigma_0) = I(\pi_0,\sigma_0) = \sup_{G^2(\sigma)\leqq\theta} I(\pi_0,\sigma),$$

とするとき，ある条件のもとで次の式を満たすplayer 1 の strategy $\overline{\pi}$ と $u^* \in R^n$ が存在する：

$$I(\pi_0, \sigma_0) = \inf_{G^1(\pi) \leqq \theta} I(\pi, \sigma_0)$$

$$= I(\overline{\pi}, \sigma_0) + \langle u^*, G^1(\overline{\pi}) \rangle.$$

亦，次の式を満たすplayer 2 の strategy $\overline{\sigma}$ と $v^* \in R^m$ が存在する：

$$I(\pi_0, \sigma_0) = \sup_{G^2(\sigma) \leqq \theta} I(\pi_0, \sigma)$$

$$= I(\pi_0, \overline{\sigma}) - \langle v^*, G^2(\overline{\sigma}) \rangle.$$

更に，ゲームの値について次の関係が成立する：

$$I(\pi_0, \sigma_0) = V^+(u^*, v^*) = V^-(u^*, v^*).$$

参考文献

1. Liu Zhaohua & K.Tanaka, On the closest multistrategy to the shadow minimum of a Markov game, 数理解析研究所講究録, 611, 1987.
2. Liu Zhaohua & K.Tanaka, On an optimal multistrategy and a weak optimal multistrategy of a Markov game, Sci. Rep. Niigata Univ., Vol.23,1987, pp 1-11.
3. K.Tanaka & C.Matsuda, On a continuously discounted vector valued Markov decision process, J.Information & Optimization Science, Vol.11,No.1,1990,pp 33-48.
4. K.Tanaka, On a discounted dynamic programming with the constraints, to appear in JMAA.
5. K.Tanaka & K.Yokoyama, On ε-equilibrium point in a noncooperative n-person game. to appear in JMAA.